# Relating morphologic and RAPD marker variation to collection site environment in wild populations of red clover (*Trifolium pratense* L.)

S.L. Greene[1], M. Gritsenko[2] and G. Vandemark[2]
[1]*USDA, ARS National Temperate Forage Legume Germplasm Resources Unit, 24106 N. Bunn Road, Prosser, WA 99350, USA;* [2]*USDA, ARS Vegetable and Forage Crop Research Unit, 24106 N. Bunn Road, Prosser, WA 99350, USA*

## Abstract

Although genotypic and phenotypic markers are used to describe genetic diversity, describing patterns of variation attributable to geographic differentiation is complex. We examined concordance between morphologic and RAPD marker classification of 33 wild red clover populations collected from the Caucasus Mountains, Russia and compared how morphologic and RAPD markers differed in their correspondence to collection site attributes. We also examined if wild red clover populations collected from sites located in areas more conducive to gene flow (i.e. adjacent to roads, or drainage systems) had the same concordance between morphologic and RAPD markers as populations collected from sites less conducive to gene flow. We measured 15 morphologic traits in a common garden and carried out a Random Amplified Polymorphic DNA (RAPD) analysis. There was a significant difference among the 33 populations for 14 out of 15 morphological traits. Morphology clustered the populations into classes that corresponded to three climate regimes. Classification schemes generated by morphologic and RAPD data did not coincide. Morphologic data corresponded with site data for populations collected at all sites. RAPD data corresponded to site data for only those populations collected at sites not conducive to gene flow. A population's adaptation to collection site needs to be considered in using neutral markers to effectively discriminating geographic differentiation. We discuss the practical lessons of this study on the effective collection, conservation and use of plant genetic resources.

*Abbreviations:* GIS – Geographic Information System, PCA – Principle Component Analysis, RAPD – Random Amplified Polymorphic DNA, NPGS – USDA National Plant Germplasm System, VIR – N.I. Vavilov Institute

## Introduction

Red clover (*Trifolium pratense* L.) is second to alfalfa in economic importance among forage legumes, and is grown for hay, silage, forage and as a soil conditioner. The species has naturalized in most temperate areas of the world. Wild-type populations (i.e. nondomesticated) persist in remote areas such as the Caucasus Mountains in southwestern Russia. Natural habitats are forest margins, meadows and cultivated field borders. Red clover is a short-lived herbaceous self-incompatible perennial commonly cross-pollinated by bumble bees (*Bombus* spp.) and honey bees

(*Apis mellifera* L.). In the wild, the main units of dispersal are the dried floret and seed. Seeds are dispersed closely to parent plants or by foraging animals.

Red clover is adapted to a wide range of pH levels, soil types and environmental conditions (Smith et al. 1985). Current interest in sustainable agriculture and natural medicinal products promises to broaden the use of red clover (Morris and Greene 2001). Broadening use emphasizes the need to exploit a wider range of diversity in red clover. Of primary importance is the need to understand and exploit genetic variation that is associated with adaptation to abiotic and biotic

factors (i.e. adaptive genetic differentiation) (Godt and Hamrick 1996). However, Godt and Hamrick (1996) acknowledge the difficulty of obtaining information on adaptive variation and concede that assessing neutral variation is frequently more efficient. In discussing local differentiation among plant populations, Linhart and Grant (1996) note that while some studies have shown concordance among morphological and molecular markers, other studies have not. They suggest that different results are to be expected depending upon whether the traits measured are selectively neutral or selectively responsive. Semagn et al. (2000) notes that the usefulness of neutral markers to characterize adaptive variation continues to be unresolved.

Several factors contribute to the complexity of the problem. Environment plays a major role in shaping the genetic structure of populations by influencing natural selection and gene flow (Linhart and Grant 1996). Greene and Hart (1999) reviewed parameters that promote ecogeographic differentiation within plant species. However, additional factors such as reproductive biology, genetic drift and migration also influence population genetic structure (Hedrick 1986; Loveless and Hamrick 1984; Slatkin 1987). The occurrence of migration through seed dispersal is rarely addressed, yet may be a significant factor contributing to the population structure of plants obtained adjacent to natural dispersal routes such as road and trail sides, or drainages. This is particularly relevant to users of ex situ collections, who frequently assume collected germplasm reflects adaptation to the site it was collected from. They frequently use collection site conditions to infer germplasm adaptation. However, many collections of wild crop genetic resources are made along easily accessed roadsides, trails, and drainage systems, locations where the genetic structure (and level of adaptation) of populations may be especially dynamic considering the high frequency of gene flow.

The objectives of this study were i) examine concordance between morphologic and RAPD marker classification of 33 wild red clover populations collected from diverse habitats; ii) examine how morphologic and RAPD markers differed in their correspondence to collection site attributes; and iii) examine if wild red clover populations collected from sites located in areas more conducive to gene flow had the same concordance between morphologic and RAPD markers as populations collected from sites less conducive to gene flow.

## Materials and methods

A joint expedition between the United States Department of Agriculture, National Plant Germplasm System (NPGS), and the Vavilov Institute of Plant Industry (VIR) was carried out in 1995 to collect forage legume species. The collection region covered an area that extended 600 km east-west and 250 km north-south, between 43–45° N latitude and 38–42° E longitude in the western Caucasus Mountains in southern Russia. Prior to the collecting trip, a GIS database was developed and maps were produced to support germplasm collection (Greene et al. 1999a). During the collection trip, the maps were used to guide the sampling of wild forms of cultivated forage species across a broad range of habitats. After the trip, individual collection sites were described using GIS attributes such as elevation, slope, and aspect derived from a digital elevation model and absolute minimum monthly temperature and monthly-accumulated rainfall derived from an interpolated climate surface (Greene et al. 1999b).

### Plant material and collection sites

Wild red clover (*Trifolium pratense* L.) populations were collected during the Caucasus Mountain exploration in 1995 in areas that had not been cultivated or actively managed for pasture improvement. Seed was collected from fifty or more randomly sampled individual plants in each population. Populations were sampled from 33 locations (Table 1). Previously, Greene et al. (1999b) classified the red clover populations into two groups using local collecting site data. The first group, designated as NONRESIDENT, represented clover populations that might not be fully adapted to site habitat since local site data suggested population establishment could have been relatively recent. This included accessions collected along roadsides, streams, construction areas, and overgrazed meadows. The 19 RESIDENT populations had been collected at sites not adjacent to obvious dispersal routes, and had no evidence of recent demographic disturbance (anthropogenic or natural). All sites were in areas not considered microhabitats (i.e. sites were assumed to be influenced by the climatic regimes described by the GIS database).

### Germplasm evaluation

In 1998, original seed of the 33 Caucasus red clover

*Table 1.* Geographic location and secondarily-derived ecogeographic attributes of 33 collection sites in the Caucasus Mountains, Russia, where wild red clover populations were collected on a joint USDA-VIR plant exploration and location and attributes of evaluation site at Prosser, WA, USA.

| PI Number | Longitude | Latitude | Elevation (m) | Slope (°) | Aspect | Annual accumulated rainfall (mm) | Annual average absolute minimal monthly temperature | Probable gene flow based on topography* |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Ecogeographic Attributes | | |
| 604739 | 38.57 | 44.78 | 178 | 5 | SE | 798 | −12.25 | + |
| 604716 | 37.55 | 44.87 | 124 | 2 | NE | 667 | −9.42 | − |
| 604715 | 41.25 | 44.30 | 834 | 3 | NE | 776 | −16.08 | − |
| 604719 | 41.43 | 44.42 | 563 | 4 | SE | 710 | −13.58 | − |
| 604693 | 43.51 | 43.72 | 488 | 0 | Flat | 538 | −10.83 | − |
| 604750 | 43.21 | 43.59 | 769 | 8 | S | 562 | −12.75 | − |
| 604709 | 40.84 | 44.74 | 281 | 2 | Flat | 684 | −12.41 | − |
| 604713 | 40.38 | 44.37 | 525 | 4 | SE | 720 | −14.33 | + |
| 604712 | 40.23 | 44.43 | 388 | 14 | SW | 761 | −13.91 | + |
| 604695 | 39.66 | 44.41 | 284 | 9 | W | 1273 | −12.83 | − |
| 604734 | 38.56 | 44.79 | 247 | 4 | NE | 803 | −12.41 | + |
| 604740 | 37.96 | 44.91 | 39 | 6 | SE | 726 | −12.91 | + |
| 604710 | 37.95 | 44.68 | 469 | 20 | W | 725 | −8.75 | + |
| 604706 | 38.35 | 44.54 | 394 | 16 | SW | 877 | −10.25 | − |
| 604723 | 39.40 | 43.85 | 48 | 13 | W | 1714 | −3.25 | + |
| 604773 | 39.67 | 43.72 | 94 | 11 | E | 1943 | −2.08 | + |
| 604751 | 40.01 | 44.09 | 1774 | 9 | NE | 1395 | −18.92 | − |
| 604704 | 40.01 | 44.06 | 1825 | 5 | W | 1469 | −18.83 | + |
| 604699 | 40.03 | 44.16 | 1281 | 1 | N | 1182 | −17.16 | + |
| 604698 | 40.25 | 44.20 | 560 | 3 | NE | 790 | −14.58 | − |
| 604771 | 40.49 | 44.62 | 305 | 8 | S | 721 | −12.58 | + |
| 604697 | 40.84 | 44.08 | 659 | 8 | W | 860 | −14.92 | + |
| 604741 | 41.60 | 43.72 | 1207 | 3 | NE | 1415 | −13.83 | + |
| 604728 | 41.68 | 43.47 | 1864 | 13 | S | 1716 | −15.42 | + |
| 604758 | 41.83 | 43.25 | 2113 | 27 | SW | 1833 | −15.25 | − |
| 604738 | 39.94 | 44.48 | 408 | 2 | NE | 959 | −14.58 | + |
| 604737 | 39.95 | 44.55 | 368 | 3 | SW | 905 | −14.83 | + |
| 604766 | 40.09 | 44.26 | 913 | 8 | SE | 970 | −16.33 | − |
| 604736 | 40.03 | 44.05 | 1868 | 17 | NE | 1494 | −19.08 | + |
| 604759 | 39.90 | 44.35 | 451 | 2 | NW | 1076 | −13.75 | + |
| 604735 | 40.18 | 43.70 | 983 | 29 | SW | 2620 | −10.25 | + |
| 604753 | 40.22 | 43.64 | 1655 | 27 | SW | 2044 | −13.66 | + |
| 604722 | 39.84 | 44.05 | 1512 | 28 | SW | 1742 | −16.00 | − |
| Prosser, WA | 46.29 | 119.73 | 112.6 | 0 | Flat | 186.2 | −6.10 | |

* (+) Populations potentially adapted to surrounding environs since local site data indicated lack of obvious seed and pollen dispersal routes (RESIDENT); (−) Populations potentially unadapted to surrounding environs since local site data indicated obvious dispersal routes such as roadways or water drainages (NONRESIDENT).

populations were planted in a field at Prosser, Washington, USA (46.29 latitude, 119.73 longitude, elevation 112.6 m). Supplemental irrigation was applied to maintain soil moisture at field capacity through the growing season. Weeds were controlled by tillage and herbicides. The common garden experiment was arranged as a randomized complete block design with four replications. For each entry, a total of 60 plants were evaluated for 15 traits (Table 2). With

the exception of leaf mark, which was measured using a six-class nominal scale, all attributes were quantitative.

In addition to the morphologic data, RAPD markers were obtained from DNA isolated from the first true set of leaves on greenhouse grown plants. 100 mg of leaf tissue per individual plant was harvested and combined to form a bulk leaf sample that represented a random sample of 18 individual plants. The leaf

*Table 2.* Minimum, maximum and mean morphological values, standard deviation, and ANOVA F value for 33 wild red clover populations evaluated at Prosser, Washington, USA in 1998 and 1999.

| Attribute | Min. | Max. | Mean | Stand. Dev. | F |
|---|---|---|---|---|---|
| Plant Height | 5.9 | 56.8 | 26.5 | 10.4 | 18.67** |
| Growth Habit | 2.0 | 5.0 | 3.6 | 0.80 | 5.16** |
| % Winter Survival | 53.3 | 100 | 85.8 | 12.8 | 1.26 |
| Leaf Mark | 1.0 | 5.6 | 2.7 | 0.73 | 5.12** |
| % Plants with 1st year bloom | 0.0 | 100 | 89.6 | 21.3 | 23.67** |
| Days to 50% bloom in 2nd year | 139.4 | 187.2 | 152.3 | 8.6 | 18.14** |
| Bloom Duration | 11.2 | 63.8 | 46.4 | 9.7 | 14.55** |
| Stem Pubescence | 1.0 | 3.8 | 2.1 | 0.5 | 6.02** |
| Flower Length | 1.4 | 1.9 | 1.63 | 0.08 | 4.69** |
| Corolla Tube Length | 0.7 | 1.71 | 1.0 | 0.08 | 1.75* |
| Short Corolla Lobe Length | 0.13 | 0.36 | 0.23 | 0.03 | 3.20** |
| Long Corolla Lobe Length | 0.3 | 0.62 | 0.46 | 0.05 | 2.08** |
| Flower Corolla Tube Color | 1.0 | 32.5 | 25.1 | 5.38 | 1.34 |
| Flower Keel Color | 11.1 | 33.0 | 29.6 | 3.2 | 1.94** |
| Flower Standard Color | 11.6 | 33.0 | 29.6 | 3.2 | 1.94** |

**, * indicates significance at P = 0.01 and 0.05 probability levels, respectively.

material was rinsed in sterile double distilled water and blotted dry prior to being frozen in liquid $N_2$. DNA was extracted by the method of Doyle and Doyle (1989). RNA was eliminated by the addition of 5U RNAase (RNAase+, $5' \rightarrow 3'$, Inc., Boulder, CO, USA) per sample. DNA was resuspended in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and the sample concentrations were determined spectrophotometrically. Samples were diluted in TE to achieve a final concentration of 20 ng/µL.

A total of 10 random 10-mer primers (Operon Technologies, Alameda, CA, USA) were used for RAPD analysis. A series of polymerase chain reactions (PCR) were conducted with 25 µL volume reactions containing 50 ng genomic DNA, 15 ng primer, 200 µM of each dNTP, and 2.5 U of Amplitaq Gold DNA polymerase (Perkin-Elmer, Norwalk, CT, USA) in buffer (50 mM KCl, 10 mM Tris- HCl, pH 8.3, 1.5 mM $MgCl_2$, and 0.001% (w/v) gelatin). Amplifications were done using a GeneAmp PCR System 9700 thermocycler (Perkin-Elmer, Norwalk, CT, USA). Each reaction was performed using an initial 'hot start' of 94 °C for 9 min, followed by 40 cycles of: 1 min at 94 °C, 1 min at 37 °C, and 1 min at 72 °C. The reactions were terminated with a final extension at 72 °C for 7 min. Amplification products were resolved on 1X TBE gels containing 1.4% agarose, stained with ethidium bromide and visualized with a source of UV light. Each primer-plant sample combination was repeated at least twice and congruence between replicates verified. Negative control reactions lacking DNA were also included for each primer. All gels were scored for both polymorphic and monomorphic bands. It was assumed that bands of the same molecular weight in different individuals were identical. Bands were scored as present or absent.

## Statistical Analysis

### Morphologic traits

To develop a set of uncorrelated morphologic attributes to minimize bias in distance estimates, a univariant analyses of variance was carried out to identify traits that did not significantly differ among accessions (Statistica, Tulsa, OK, USA). Nonsignificant attributes were excluded from further analysis. A Pearson product-moment correlation matrix was calculated for the remaining attributes to identify significantly correlated attributes. If attributes were significantly correlated with an *r* value > 0.6, only one attribute was included for cluster analysis. Johnson (1998) suggested that attributes having an *r* value > 0.60 have a strong enough linear relationship that one attribute can be used as a surrogate for the other. By including only one attribute from correlated attributes, bias was minimized in calculating resemblance coefficients due to the weighting effect of redundant attributes (Romesburg 1990). NTSYS-pc program version 2.02k (Rohlf 1993) was used for the subsequent analysis. Attributes were standardized to a mean of zero and a variance of one and Euclidean distance coefficients calculated to generate three morphologic distance matrices (ALL, NONRESIDENT,

RESIDENT). The ALL distance matrix was used to produce a dendrogram based on the unweighted pair-group method with arithmetic averages (UPGMA). To verify the cluster analysis and to assist in visualizing the data (Johnson 1998), principal component analysis (PCA) was carried out. The ALL distance matrix was transformed to scalar product form using the double-centered TRANSFORMATION module, so eigenvalues and eigenvectors could be computed. Accessions were plotted with the first and second eigenvectors using the ORDINATION module of NTSYS-pc.

*RAPD markers*
The RAPD data were used to construct a pairwise genetic distance matrix. Since red clover is a perennial out-crossing species, we used the procedure by Nebauer et al. (1999) to identify polymorphic RAPDS bands that had a very low frequency of recessive alleles ($q^2 < 3/n$, where $q^2$ = frequency of bands scored 0 and n = total number of bands). These bands, along with monomorphic bands were dropped from the analysis, leaving a set of uncorrelated polymorphic RAPD bands that were subsequently used to estimate pairwise genetic distance to minimize bias in the calculation of genetic distance using RAPD markers (Lynch and Milligan 1994).

The correspondence coefficient used was the simple matching coefficient (SM). The SM is a similarity coefficient, so to simplify the correlation of RAPD matrices with the morphological and GIS-derived matrices, which were based on dissimilarity coefficients, we converted the RAPD matrix from a similarity to a dissimilarity matrix by taking 1-s, where s = SM. The RAPD data were used to generate three genetic distance matrices (ALL, RESIDENT and NONRESIDENT). Dendrograms and scatter plots were generated using the same methods used with the morphologic data.

*Ecogeographic attributes*
A geographic distance matrix based on the latitude and longitude of collection sites was calculated. Pair wise distance matrices were also developed for eight sets of GIS-derived attributes whose combined effect might influence population genetic structure: 1) elevation, slope, aspect, moisture zone and temperature zone; 2) slope, aspect, moisture zone, temperature zone; 3) elevation, slope and aspect; 4) moisture zone and temperature zone; 5) moisture zone, temperature zone, elevation; 6) elevation; 7) elevation and

latitude; and 8) moisture zone, temperature zone and latitude. The comparison between the ecogeographic matrices with the RAPD and morphological matrices was carried out separately for NONRESIDENT and RESIDENT populations.

Attributes were standardized to a mean of zero and a variance of one and Euclidean distance coefficients were calculated to generate a set of distance matrices. The same methods used with the morphological data and RAPD data were used to generate dendrograms and scatter plots of the GIS-derived data.

*Matrix correlation*
Correlation between distance matrices generated from morphologic, RAPD marker and GIS-derived data was compared using the product moment correlations (*r*) derived from the normalized Mantel Z (Mantel 1967). Matrices comparisons were carried out using the MXCOMP module of the NTSYS-pc program version 2.02k (Rohlf 1993). The estimated *Z* test criterion was compared to the randomized distribution of *Z* obtained from 1000 random permutations of the matrices (excluding the observed comparison) to determine the probability of obtaining a random *Z* greater than the estimated *Z* (Manly 1994). Correspondence between the morphologic and RAPD distance matrices, and each of these with geographic distance matrices were examined. The eight different GIS data matrices were compared against the matrices from morphologic and RAPD marker data for both RESIDENT and NONRESIDENT accessions. To control for Type I error rate in families of comparisons, we used Benjamini and Hochberg (1995) multiple comparison procedure for controlling false discovery rate.

**Results**

The 33 Caucasus collection sites reflected 25 different climate regimes based on timing of rainfall throughout the year, and seasonal absolute minimum temperatures (Figure 1). At Prosser, WA, temperatures were within the same range as seen in the Caucasus Mountains, although average rainfall was considerably less (Fig. 1(o)). Supplemental irrigation compensated for this, and provided optimal soil moisture for the common garden trial.

*Morphologic classification*
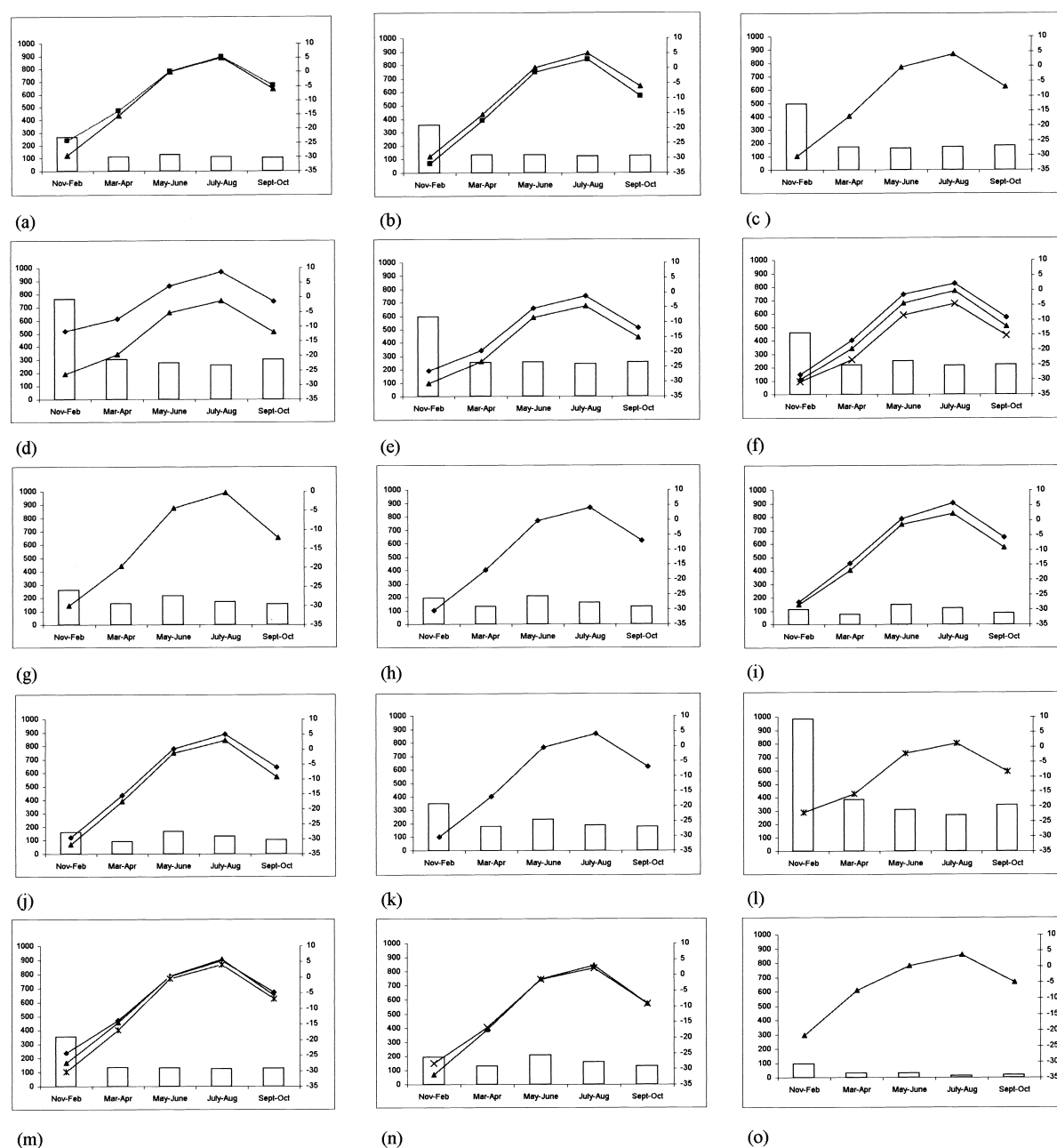
There were significant differences among the 33 wild

*Figure 1.* Climatographs indicating seasonal rainfall (left axis (mm) and bars) and absolute minimum monthly temperatures (right axis (° C) and lines) at the sites where wild red clover populations were collected in the Caucasus Mountains, Russia, and the evaluation site at Prosser, WA, USA. 33 red clover accessions were collected from a total of 25 distinct environments.(a) □ PI 604716, 604710, △ PI 604740; (b) □ PI 604706, △ PI 604734; (c) PI 604695; (d) ◇ PI 604723, 604773, △ PI 604758, 604753, 604722; (e). ◇ PI 604728, △ PI 604704, 604736; (f). ◇ PI 604699, △ PI 604741, ☆ PI 604751; (g). PI 604766; (h). PI 604712; (i) △ PI 604750, ◇ 604693; (j) ◇ PI 604709, 604771; △ PI 604719, 604713; (k) PI 604759; (l) PI 604735; (m) ◇ PI 604706, △ PI 604739, ☆ PI 604738; (n) △ PI 604715, 604697; ☆ PI 604698; (o) Prosser, Washington, USA.

populations for 13 out of the 15 morphological traits examined in the evaluation trial. Table 2. lists the minimum, maximum and mean values, stardard de-viation and F value for the 15 traits. The following traits were significantly correlated: plant height and growth habit ($r = 0.67$); flower length and corolla

tube length ($r = 0.52$), shortest corolla lobe length and longest corolla lobe length ($r = 0.34$); flower standard color and flower keel color ($r = 0.99$). Significant morphological differences among the 33 accessions, and trait correlations suggested that the following set of attributes would contribute to an unbiased estimate of morphological distance: growth habit, flower length, pubescence, leaf mark, % flowering 1st year, winter survival, days to bloom, duration of bloom, and flower standard color. Winter survival was included because although the 33 accessions did not differ statistically, a broad range of survival (53% to 100%) was represented and the trait itself is an important plant limiting character.

Fig. 2(a) illustrates the dendrogram based on the cluster analysis of the morphological distance matrix of 33 accessions. The correlation between the distance matrix and cophenetic matrix, for the morphologic data was $r = 0.83$, indicating the dendrogram represented the distance matrix with minimal distortion for the 33 wild populations. Scatter plots of the first three principal component scores visually supported the cluster analysis. The first three dimensions of the PCA accounted for 72% of the morphologic variation. The dendrogram was cut at a distance of 4.32, which represented the center of the widest range of distance measures corresponding to the same number of clusters (Romesburg 1990) and resulted in five morphologic classes. We recognized three classes, since two of the five classes contained only one member that was similar to members of the adjacent class except for a single attribute. Seventy-five percent of the accessions fell into a class that was characterized by upright growth habit, generally good winter survival, normal or delta leaf marks, less stem pubescence, first year bloom and had flowers that were smaller with pale purplish standards. These accessions were the earliest and longest blooming populations. Accessions in this class had been collected at lower elevations (average elevation was 516 m), had significantly less rainfall throughout the year (average annual accumulation of 993 mm) and warmer temperatures throughout the growing season. Although the average annual absolute minimum was −9.5 °C, winter temperatures were comparable to the other two classes. The second class contained five accessions, characterized by the greatest level of pubescence, slightly larger and darker flowers and were medium in maturity and bloom duration. Accessions in this class had been collected at higher elevations (average elevation was 1518 m), had significantly more rainfall

throughout the year (average annual accumulation of 1440 mm) and had the coldest temperatures during the winter and growing season (average annual absolute minimum was −13.9 °C). The third class contained three accessions that had greatest winter survival at the Prosser, WA site, lowest percent of first year flowering, latest bloom date, and shortest duration of bloom of all classes. Accessions in this class had been collected at the highest elevations (average elevation was 1653 m), had the most rainfall throughout the year (average annual accumulation of 1933 mm) and had cold temperatures during the winter and growing season (average annual absolute minimum was −11.6 °C).

*RAPD marker classification*

To ensure an unbiased estimate of genetic distance among the 33 accessions using RAPD markers, 38 out of the 83 total bands were dropped since they appeared in low frequency or were monomorphic. The remaining 45 bands did not cluster the accessions into classes that were as distinct as those formed by the morphologic data (Fig. 2(b)). The cophenetic correlation was $r = 0.72$, which indicated the dendrogram produced from the cluster analysis gave a relatively poor fit when compared with the distance matrix. Scatter plots of principle component scores did not support the cluster analysis and PCA indicated that only 28.4 % of the RAPD variation was accounted for in the first three components.

There was no correspondence between the distance matrices based on morphologic data and RAPD marker data. When the distance matrices of the RESIDENT populations were compared using a Mantel test, the matrices were not significantly correlated ($r = 0.11$; $P \leq 0.82$). There was also no correspondence between morphologic and genetic distance for NON-RESIDENT populations ($r = -0.10$; $P \leq 0.25$).

*Correspondence of morphology and RAPD markers to collection sites*

Correspondence of morphologic and RAPD marker distance to ecogeographic distances was not the same, and differed for RESIDENT and NONRESIDENT populations. The morphologic distance matrices for RESIDENT and NONRESIDENT populations were significantly correlated to the distance matrices of all eight combinations of ecogeographic attributes (Table 3). Correlation was generally stronger among NON-

Euclidean Distance Coefficient

(a)



Simple Matching Coefficient
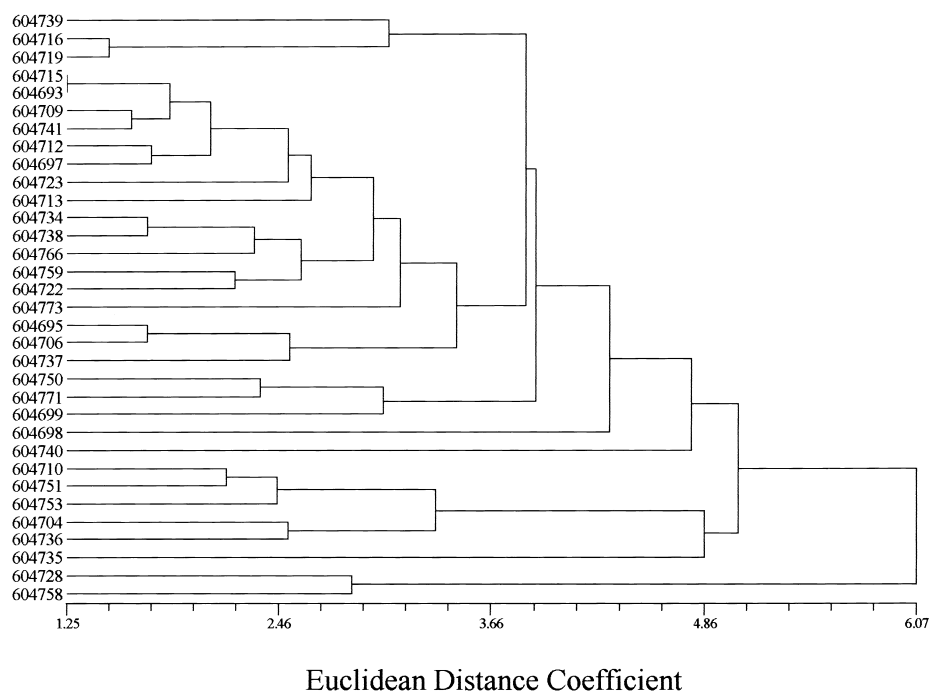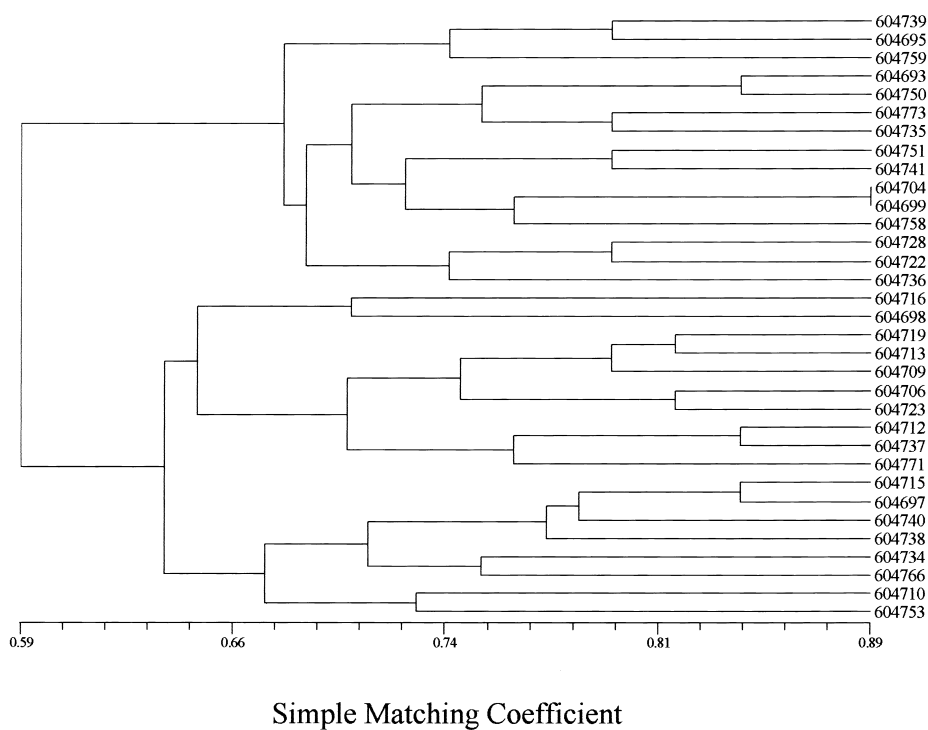
(b)

*Figure 2.* Dendrogram based on the cluster analysis of 33 populations of wild red clover. (a). Distance matrix based on morphological traits; (b). Distance matrix based on RAPD marker data.

*Table 3.* Correlation between morphologic and ecogeographic distance matrices for 19 populations of red clover classified as RESIDENT and 14 populations of red clover classified as NONRESIDENTS of the collection site.

| Collection site attributes | RESIDENT | NONRESIDENT |
|---|---|---|
| | $r^\dagger$, P | $r^\dagger$, P |
| Elevation, Slope, Aspect, Moisture Zone, Temperature Zone | 0.45, 0.002* | 0.51, 0.002* |
| Slope, Aspect, Moisture Zone, Temperature Zone | 0.31, 0.004* | 0.37, 0.005* |
| Elevation, Slope, Aspect | 0.44, 0.004* | 0.48, 0.002* |
| Moisture Zone, Temperature Zone | 0.25, 0.004* | 0.28, 0.026* |
| Moisture Zone, Temperature Zone, Elevation | 0.44, 0.002* | 0.48, 0.002* |
| Elevation | 0.55, 0.002* | 0.58, 0.002* |
| Elevation, Latitude | 0.49, 0.002* | 0.65, 0.002* |
| Moisture Zone, Temperature Zone, Latitude | 0.32, 0.002* | 0.50, 0.004* |

$^\dagger r$ = normalized Mantel Z statistic; *$P$ < P calculated for 5 % significance level using false discovery rate procedure (Benjamini and Hochberg 1995) to control Type I error; NS = not significant (significance tests after 1000 permutations).

RESIDENT populations then RESIDENT populations. The strongest correlations between morphology and collection site attributes for both RESIDENT and NONRESIDENT populations occurred with elevation and latitude, latitude alone, and elevation, slope, aspect, moisture and temperature zone. The weakest morphologic correlation for both RESISENT and NONRESIDENT populations was with the distance matrices based upon moisture and temperature zone alone.

In contrast to morphology, RESIDENT and NON-RESIDENT populations differed in how distance matrices based on RAPD markers corresponded to the ecogeographic distance matrices (Table 4). RESIDENT populations showed a significant correspondence to all ecogeographic distance matrices, except for matrices comprised of: elevation, slope, aspect, moisture and temperature zone; slope, aspect, moisture and temperature zone; and moisture and temperature zone. The strongest correlation was with the following matrices: elevation and latitude; moisture zone and temperatue zone; and elevation. Correlation among RESIDENT populations was generally weaker between RAPD marker-based distance and ecogeographic distance then compared with correlations between morphologic and ecogeograpgic distances. In contrast to the morphological data, the RAPD-derived data on the NONRESIDENT populations showed no significant correlation with any of the ecogeographic distance matrices. When geographic distance was examined, there was no correlation between geographic distance and morphologic and genetic distance for both RESIDENT and NONRESIDENT red clover populations.

## Discussion

In our study area in the Caucasus Mountains, Russia, red clover populations were found growing at elevations ranging from 40 to 2000 m, with annual precipitation ranging from 876 to 2289 mm, and in soils with soil pH ranging from 3.7 to 7.7. In a common garden experiment there were significant morphological differences among the 33 populations examined with the exception of winter hardiness. That the common

*Table 4.* Correlation between genetic and ecogeographic distance matrices for 19 populations of red clover classified as RESIDENT and 14 populations of red clover classified NONRESIDENTS of the collection site.

| Collection Site Attributes | RESIDENT | NONRESIDENT |
|---|---|---|
| | $r^\dagger$, P | $r^\dagger$, P |
| Elevation, Slope, Aspect, Moisture Zone, Temperature Zone | 0.02, 0.39 NS | 0.14, 0.13 NS |
| Slope, Aspect, Moisture Zone, Temperature Zone | 0.1, 0.88 NS | 0.06, 0.30 NS |
| Elevation, Slope, Aspect | 0.13, 0.01* | 0.18, 0.07 NS |
| Moisture zone, Temperature Zone | 0.14, 0.05 NS | 0.02, 0.41 NS |
| Moisture zone, Temperature Zone, Elevation | 0.18, 0.02* | 0.14, 0.10 NS |
| Elevation | 0.2, 0.03* | 0.28, 0.02 NS |
| Elevation, Latitude | 0.28, 0.006* | 0.29, 0.02 NS |
| Moisture Zone, Temperature Zone, Latitude | 0.24, 0.006* | 0.11, 0.17 NS |

$^\dagger r$ = normalized Mantel Z statistic; *$P$ < P calculated for 5 % significance level using false discovery rate procedure (Benjamini and Hochberg 1995) to control Type I error; NS = not significant (significance tests after 1000 permutations).

garden location had a mild winter likely accounted for this result. Morphological data separated the 33 populations into three distinct groups of germplasm that corresponded respectively to sites at a) low elevations, with a dryer warmer growing season, b) high elevation with medium levels of precipitation and cold temperatures, c) high elevations with high levels of precipitation and cold temperatures. The results suggest that the strongly heterogeneous nature of the Caucasus Mountains and corresponding strong differences in selection regimes may have overcome the effects of common ancestry and isolation by distance.

In contrast to morphology, RAPD data did not separate the accessions into distinct classes. Several factors may account for the discordance between morphologic and RAPD markers. Johns et al. (1997) suggested that discordance might occur if morphological similarity was due to different combinations of alleles producing similar phenotypes. Discordance between morphologic and molecular markers could also occur if a single or few genes controlled the expression of morphological traits, and were not detected by RAPDs (Steiner and Garcia de los Santo 2001). Discordance may also be due to differences in evolutionary rates between morphologic characters and characters originating from selectively neutral, non-coding DNA, especially if the morphological characters have adaptive value, and molecular markers reflect functional neutrality (Johns et al. 1997; Linhart and Grant 1996).

Morphological variation was strongly associated with environmental variation. Correspondence was observed for populations collected at sites located in areas less conducive to gene flow and at sites adjacent to obvious dispersal routes (i.e. areas more conducive to gene flow). Surprisingly, correlations tended to be stronger with NONRESIDENT populations (those collected at sites along obvious dispersal routes). A possible explanation may be that long-term accumulation of neutral mutations in RESIDENT populations has slightly obscured the correspondence between morphology and environment (Linhart and Grant 1996).

RAPD variation was also associated with environmental variation, but only for populations collected at geographically isolated sites (i.e. areas less conducive to gene flow). Although significant, the association was much weaker then the correspondence between morphology and environment. Heaton et al. (1999) suggested a number of reasons why there might be a lack of association between DNA molecular markers

and ecology, including the homogenizing influence of long distance gene flow, failure of primers to identify genetic differences due to limited sampling of the genome and/or sampling of neutral markers not associated with adaptation and phenotypic plasticity of the examined populations. Our results support other studies that suggest that selectively responsive markers (such as morphologic traits) may be more effective at understanding adaptive variation, then markers that are selectively neutral (Johns et al. 1997; Linhart and Grant 1996).

Although weak, the significant correlation between genetic distance and ecogeographic distance for RESIDENT populations, and non-significant correlation for NONRESIDENT populations, suggested that the RAPD markers might have some association with adaptive alleles; given the premise that RESIDENT populations might have accumulated more adaptive traits then NONRESIDENT populations. Other studies have reported a positive correspondence between RAPD markers and ecology. Li et al. (1999) identified adaptive DNA differentiation in wild emmer wheat growing in two contrasting microhabitats. Semagn et al. (2000) reported a strong association between RAPD markers and altitude, temperature and rainfall in *Phytolacca dodecandra*.

In our study, no significant correlation was found between geographic distance and either morphological distance or RAPD marker distance. Similar studies that examined correspondence between geographic distance and distance measures based upon RAPD markers have reported positive (i.e. Müller-Schärer and Fischer 2001) and negative associations (i.e. Nebauer et al. 1999). Reasons for inconsistencies may include differences in geographic scale examined, type of breeding system in studied species, and level of heterogeneity in environment examined.

Practical lessons can be drawn from this study relevant to the effective collection, conservation and use of crop genetic resources. This study supports the widely accepted strategy of sampling geographically distinct populations to ensure sampling of intraspecific genetic variation, in that both morphological and molecular marker diversity corresponded to the ecogeographic diversity of collecting sites. This study suggested that the presence of geographic features that prevent gene flow are an important factor in promoting geographic differentiation in populations. We suggest that collectors not only document the site they collect from, but also provide information on relative levels of gene flow that may occur (based on

local topographic considerations). This type of information will help us understand the potential genetic structure of accessions, allowing us to better utilize the appropriate tools to answer questions of interest. Further studies are needed to understand the best methods for characterizing patterns of genetic diversity attributable to geographic differentiation. This includes conducting experiments that use molecular markers linked to adaptive traits or that are punitive alleles for adaptive traits, comparing common garden environments that imposed similar and different selection pressure and that examine the influence of life history traits such as breeding system.

## Acknowledgements

## References

Benjamini Y. and Hochberg Y. 1995. Controlling the false discovery rate. a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57: 289–300.

Doyle J.J. and Doyle J.L. 1989. Isolation of plant DNA from fresh tissue. Focus 12: 13–15.

Godt M.W. and Hamrick J.L. 1996. Genetic diversity and morphological differentiation in *Liatri helleri* (Asteraceae), a threatened plant species. Biodiversity and Conservation 5: 461–471.

Greene S.L., Hart T.C. and Afonin A. 1999a. Using geographic information to acquire wild crop germplasm for *ex situ* collections. I. Map Development and Use. Crop Science 39: 836–842.

Greene S.L., Hart T.C. and Afonin A. 1999b. Using geographic information to acquire wild crop germplasm for *ex situ* collections. II. Post collection analysis. Crop Science 39: 843–849.

Greene S.L. and Hart T.C. 1999. Implementing a geographic analysis in germplasm conservation. In: Greene S.L. and Guarino L. (eds), Linking Genetic Resources to Geography. Strategies for Conserving Crop Biodiversity. CSSA, Madison, WI, Special Publication No 27, pp. 25–38.

Heaton H.J., Whitkus R.R. and Gomez-Pompa A. 1999. Extreme ecological and phenotypic differences in the tropical tree chicozapote (*Manilkara zapota* (L.) P. Royen) are not matched by genetic divergence. a random amplified polymorphic DNA (RAPD) analysis. Molecular Ecology 8: 627–632.

Hedrick P.W. 1986. Genetic polymorphisms in heterogeneous environments. a decade later. Annual Review of Ecological Systematics 17: 535–566.

Johns M.A., Skroch P.W., Nienhuis J., Hinrichsen P., Bascur G. and Munoz-Schick C. 1997. Gene pool classification of common bean landraces from Chile based on RAPD and morphological data. Crop Science 37: 605–613.

Johnson D.E. 1998. Applied Multivariate Methods for Data Analysts. Brooks/Cole Publishing Company, CA.

Jordano P. and Godoy A. 2000. RAPD variation and population genetic structure in *Prunus mahaleb* (Rosaceae), an animal-dispersed tree. Molecular Ecology 9: 1293–1305.

Linhart Y.B. and Grant M.C. 1996. Evolutionary significance of local genetic differentiation in plants. Annual Review of Ecological Systematics 27: 237–277.

Li Y.C., Fahima T., Beiles A., Korel A.B. and Nevo E. 1999. Microclimatic stress and adaptive DNA differentiation in wild emmer wheat, *Triticum dicoccoides*. Theoretical and Applied Genetics 98: 873–883.

Loveless and Hamrick 1984. Ecological determinants of genetic structure in plant populations. Ann. Rev. Ecol. Syst. 15: 65–95.

Lynch M. and Milligan B.G. 1994. Analysis of population genetic structure with RAPD markers. Molecular Ecology 3: 91–99.

Manly B.F.J. 1994. Multivariate Statistical Methods. a primer. 2nd edn. Chapman and Hall, New York.

Mantel N.A. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209–220.

Morris J.B. and Greene S.L. 2001. Defining a multiple-use germplasm collection for the genus *Trifolium*. Crop Science 41: 893–901.

Müller-Schärer H. and Fischer M. 2001. Genetic structure of the annual weed *Senecio vulgaris* in relation to habitat type and population size. Molecular Ecology 10: 17–28.

Nebauer S.G., Del Castillo-Agudo L. and Segura J. 1999. RAPD variation within and among natural populations of outcrossing willow-leaved foxglove (*Digitalis obscura* L.). Theoretical and Applied Genetics 98: 985–994.

Rohlf F.J. 1993. NTSYS: Numerical Taxonomy and Multivariate Analysis System, v. 2.02k. Exeter Software, Setauket, New York.

Romesburg H.C. 1990. Cluster Analysis for Researchers. Krieger Publishing Company, Malabar, Florida.

Semagn K., Bjornstad A., Stedje B. and Bekele E E. 2000. Comparison of multivariate methods for the analysis of genetic resources and adaptation in *Phytolacca dodecandra* using RAPD. Theoretical and Applied Genetics 101: 1145–1154.

Slatkin 1987. Gene flow and the geographic structure of natural populations. Science 236: 787–792.

Smith R.R., Taylor N.L. and Bowley S.R. 1985. Red Clover. In: Taylor N.L. (ed.), Clover Science and Technology. ASA Special Publication 25, Madison, WI, pp. 471–490.

Steiner J.J., Beuselinck P.R., Greene S.L., Kamm J.A., Kirkbride J.H. and Roberts C.A. 2001. A description and interpretation of the NPGS Trefoil core subset collection. Crop Science 41: 1968–1980.

Steiner J.J. and Garcia de los Santo G. 2001. Adaptive ecology of *Lotus corniculatus* L. genotypes. I. Plant morphology and RAPD marker characterizations. Crop Science 41: 552–563.